

DEVICE FOR DETERMINING TONE OF SKIN IN VIDEO SIGNAL

Publication number: JP10162128

Publication date: 1998-06-19

Inventor: OKADA HIROYUKI; ROSENBERG JONATHAN DAVID

Applicant: LUCENT TECHNOLOGIES INC; SHARP KK

Classification:

- international: H04N9/64; G06K9/00; G06T1/00; G06T7/00; H04N7/26; H04N7/36; H04N7/50; H04N9/64; G06K9/00; G06T1/00; G06T7/00; H04N7/26; H04N7/36; H04N7/50; (IPC1-7): G06T1/00; G06T7/00; H04N9/64

- European: G06K9/00F2; H04N7/26J2; H04N7/36D; H04N7/50E5H

Application number: JP19970275963 19971008

Priority number(s): US19960727862 19961008

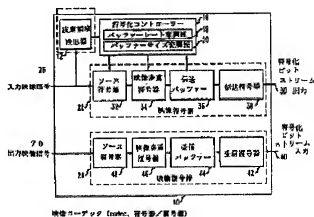
Also published as:

EP0836326 (A2)
US6343141 (B1)
EP0836326 (A3)

Report a data error here

Abstract of JP10162128

PROBLEM TO BE SOLVED: To make the hand, face, and other skin areas of a person present in a scene recognizable, by providing a locator, which analyzes a video signal and recognizes an object in a desired shape and a detector, which analyzes pixels from the object. **SOLUTION:** A skin area detector 12 is arranged in a video CODEC 10 and includes the shape locator and the tone detector. The shape locator searches for a part as a skin area in a video frame through the recognition of all object edges in the video frame and a judgment of whether or not an arbitrary edge is close to the outline of the specific shape. A recognized object which is possibly the skin area is analyzed by the tone detector, and it is judged whether or not pixels of the object have featured signal energy in the skin area. This skin area detector 12 is interposed between an input video signal 26 and an encoding controller 16 and provides input which is relative to the position of the skin area in a video sequence prior to encoding.



Data supplied from the esp@cenet database - Worldwide

Family list5 family members for: **JP10162128**

Derived from 3 applications

[Back to JP10162128](#)**1 Skin area detection for video image systems****Inventor:** OKADA HIROYUKI (US); ROSENBERG JONATHAN DAVID (US)**EC:** G06K9/00F2; H04N7/26J2; (+2)**Publication info:** **EP0836326 A2** - 1998-04-15
EP0836326 A3 - 2004-03-31**Applicant:** LUCENT TECHNOLOGIES INC (US); SHARP KK (JP)**IPC:** *H04N9/64; G06K9/00; G06T1/00* (+13)**2 DEVICE FOR DETERMINING TONE OF SKIN IN VIDEO SIGNAL****Inventor:** OKADA HIROYUKI; ROSENBERG JONATHAN DAVID**EC:** G06K9/00F2; H04N7/26J2; (+2)**Publication info:** **JP3256474B2** - 2002-02-12
JP10162128 A - 1998-06-19**Applicant:** LUCENT TECHNOLOGIES INC; SHARP KK**IPC:** *H04N9/64; G06K9/00; G06T1/00* (+14)**3 Skin area detection for video image systems****Inventor:** OKADA HIROYUKI (US); ROSENBERG JONATHAN DAVID (US)**EC:** G06K9/00F2; H04N7/26J2; (+2)**Publication info:** **US6343141 B1** - 2002-01-29**Applicant:** LUCENT TECHNOLOGIES INC (US)**IPC:** *H04N9/64; G06K9/00; G06T1/00* (+15)

Data supplied from the **esp@cenet** database - Worldwide

特開平10-162128

(43)公開日 平成10年(1998) 6月19日

(51)Int.Cl.⁴

識別記号

F I

G 0 6 T 1/00

G 0 6 F 15/62

3 8 0

Z

7/00

H 0 4 N 9/64

3 1 0

H 0 4 N 9/64

審査請求 未請求 請求項の数22 O L (全 15 頁)

(21)出願番号 特願平9-275963

(71)出願人 59607/259

(22)出願日 平成9年(1997)10月8日

ルーセント テクノロジーズ インコーポ
レイテッド
Lucent Technologies
Inc.

(31)優先権主張番号 08/727862

アメリカ合衆国 07974 ニュージャージ
ー、マレーヒル、マウンテン アベニュー
600-700

(32)優先日 1996年10月8日

(71)出願人 00000049

(33)優先権主張国 米国 (US)

シャープ株式会社
大阪府大阪市阿倍野区長池町22番2号

(74)代理人 弁理士 三俣 弘文

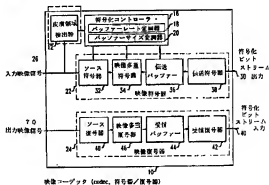
最終頁に続く

(54)【発明の名称】 映像信号における皮膚のトーンを定める装置

(57)【要約】

【課題】 映像システムにおいて皮膚領域を検出する。

【解決手段】 映像系列における皮膚領域を検出する装置であり、この装置は、形状ロケータ（形状位置決め器）及び色調検出器を含む。形状ロケータは、入力映像系列を解析し、映像フレームにおける全対象（物）のエッジを認識し、当該エッジが、皮膚領域を含む可能性のある所定の形状の輪郭に近似するか判断する。皮膚領域を含む得る対象の位置が定められると、色調検出器は、当該対象の画素を検証し、皮膚領域に特徴的な信号エネルギーを有するか判断し、さらに、そのような信号エネルギーをもつ画素をサンプリングし、皮膚の色調の範囲を決定し、全フレームにおける色調と比較し、適合する皮膚の色調すべてを見つ出す。



【特許請求の範囲】

【請求項1】 映像信号を解析し、望みの形状を有する対象を認識するローケータと、

解析した画素が皮膚領域を示す輝度パラメータを有しているかを判断するために、前記望みの形状を有する、認識された前記対象からの画素を解析する検出器とを有することを特徴とする前記映像信号における皮膚のトーンを定める装置。

【請求項2】 前記望みの形状は、通常、皮膚領域を含む形状であることを特徴とする請求項1の装置。

【請求項3】 前記望みの形状は、人間の形状と関連した弧形を有することを特徴とする請求項2の装置。

【請求項4】 前記望みの形状は、楕円形であることを特徴とする請求項3の装置。

【請求項5】 前記皮膚領域を示す前記輝度パラメータは、解析された前記画素についての交流(AC)信号エネルギー成分であることを特徴とする請求項1の装置。

【請求項6】 前記検出器が、画素の色パラメータを定めるために、解析された該画素をさらにサンプリングすることを特徴とする請求項1の装置。

【請求項7】 前記色パラメータは、クロミナンスパラメータであることを特徴とする請求項6の装置。

【請求項8】 前記検出器が、さらに比較器を有し、この比較器は、解析された前記画素について定められた前記色パラメータと同一となっている、未解析の画素における複数の色パラメータを認識するために、解析された前記画素について定められた前記色パラメータを、前記映像信号についての未解析の画素における該複数の色パラメータと比較することを特徴とする請求項6の装置。

【請求項9】 解析された前記画素についての前記色パラメータの位置を基に、符号器が符号セグメントを生成することを特徴とする請求項1の装置。

【請求項10】 前記形状が目一鼻孔領域と関連した画素を含んでいるかを判断するために、前記人間の形状と関連した前記弧形について解析がなされることを特徴とする請求項3の装置。

【請求項11】 前記目一鼻孔領域と関連のない前記画素については、前記検出器により解析が省かれることを特徴とする請求項10の装置。

【請求項12】 望みの形状を有する対象を認識するために、映像信号を解析するステップと、
解析された前記画素が皮膚領域を示す輝度パラメータを有しているかを判断するために、前記望みの形状を有する、認識された前記対象からの画素を解析するステップと、を有することを特徴とする前記映像信号における皮膚のトーンを定める方法。

【請求項13】 前記望みの形状は、通常、皮膚領域を含む形状であることを特徴とする請求項12の方法。

【請求項14】 前記望みの形状は、人間の形状と関連した弧形を有することを特徴とする請求項13の方法。

【請求項15】 前記望みの形状は、楕円形であることを特徴とする請求項14の方法。

【請求項16】 前記皮膚領域を示す輝度パラメータは、解析された前記画素についての交流(AC)信号エネルギー成分であることを特徴とする請求項12の方法。

【請求項17】 画素の色パラメータを定めるために、解析された該画素をサンプリングするステップをさらに有することを特徴とする請求項12の方法。

【請求項18】 前記色パラメータは、クロミナンスパラメータであることを特徴とする請求項17の方法。

【請求項19】 解析された前記画素について定められた前記色パラメータと同一となっている、未解析の画素における複数の色パラメータを認識するために、解析された前記画素について定められた前記色パラメータを、前記映像信号についての未解析の画素における該複数の色パラメータと比較するステップをさらに有することを特徴とする請求項12の方法。

【請求項20】 解析された前記画素についての前記色パラメータの位置を基に、符号セグメントを生成するステップをさらに有することを特徴とする請求項17の方法。

【請求項21】 前記形状が目一鼻孔領域と関連した画素を含んでいるかを判断するために、前記人間の形状と関連した前記弧形を解析するステップをさらに有することを特徴とする請求項14の方法。

【請求項22】 前記目一鼻孔領域と関連のない前記画素については、前記検出器により解析が省かれることを特徴とする請求項21の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、映像を用いた電話会議システムといった、マルチメディアへの応用向けの低ビット速度通信システムに関するものである。より特定すると、映像における皮膚領域の認識を行う方法及び、システムに関するものである。

【0002】

【従来の技術】フルカラー、フルモーションの映像の保存及び伝送への需要は増加しつつある。これらの映像は、映画やテレビの製作におけるようなエンターテインメント(娯楽)向けのみならず、技術的解析及び医療用画像といった解析及び診断業務にも用いられている。

【0003】これらの映像をデジタル形式で提供することには、幾つもの利点がある。例えば、デジタル画像は、画質の向上(強調)や操作がより行い易いのである。また、デジタル映像の場合、最小限の信号劣化で、何回にもわたって精緻に再生がなれることが可能であ

る。

【0004】一方、デジタル映像は、保存に相当量のメモリ容量を必要とし、同様に、伝送に高バンド幅（高帯域幅）のチャネルを必要とする。例えば、256段階のグレーレベル（濃度レベル）を有する、512×512画素のグレースケール（無彩色スケール）での単一の画像は、その保存に256,000バイト以上を必要とする。フルカラーの画像では、ほぼ800,000バイトを必要とする。動きが自然に見えるには、毎秒少なくとも30回は画像が更新されることを要する。

【0005】従って、動きが自然に見える、フルカラーの動画像向けの伝送チャネルは、毎秒およそ190メガビット提供するものでなくてはならない。しかしながら、テレビ電話、ビデオオンデマンド用セットトップボックス、映像による電話会議システムを含めた、今日のデジタル通信の応用例では、伝送チャネルにバンド幅（帯域幅）の限界があり、映像情報を伝送するのに利用可能なビット数は、毎秒190メガビット以下である。

【0006】その結果、デジタル映像信号を保存及び伝送するに必要とされる情報量を低減させるのに、例えば、離散コサイン変換（discrete cosine transformation, DCT）といった、幾多もの画像圧縮技術が用いられてきた。これらの技術は一般に、デジタル映像を伝送、記録、複製するために用いられたデータ量を低減させるべく、本来の画像に存する多量の冗長性を利用するものである。

【0007】例えば、伝送される画像が、晴天の空の画像であるとするれば、離散コサイン変換（DCT）による画像データ情報は、多くのゼロデータ成分を有していることとなる。というのは、そのような画像で描写された対象（物）においては、変化がごく僅かか、あるいはほとんど存在していないからである。そこで、晴天の空の画像情報は、少数の非ゼロデータ成分のみを伝送することで圧縮されるのである。

【0008】離散コサイン変換（DCT）のような画像圧縮技術に関連した、一つの問題点は、このような技術では、損失画像が生じるということである。というのは、ビット速度を低減させるため、部分的な画像情報のみが伝送されるからである。損失画像とは、復号された対象内容を元の画像内容と比較すると、描かれた対象（物）にひずみを含んでいる映像のことである。

【0009】映像による電話会議や電話による応用例のほとんどの場合、背景よりもむしろ人を含んでいる画像へ関心が向けられていることから、ひずみのない映像を伝送する能力が重要視されるのである。これは、見る者（ビューアー）は自らの注意を、例えば衣服や背景といったものに対する代わりに、映像シーケンスに含まれた、当該場面に存する人の顔、手、あるいはその他の皮膚領域といった特定の特徴（対象）へ向ける傾向があると考えられるからである。

【0010】幾つかの状況では、映像シーケンスに含まれた顔の特徴を極めてうまく表現するということは、その理解しやすさにとって優れたものとなるのである。例えば、読者術に依存していることが考えられる聴力障害のあるビューアーの場合が、そのような例である。そのような応用例の場合、顔の領域についてひずみを受けたものを含んでいる、復号された映像シーケンスは、見る者（ビューアー）にとって困ったものとなりうる。

【0011】というのは、そのような画像シーケンスは、顔の特徴が過度に平滑化されて描かれている場合が多く、顔に人工的な特質を与えているのである。例えば、元の映像に見られる顔に存するしわのような、細かい顔の特徴は、圧縮され伝送された映像を復号したものにおいては消われている傾向があり、このようなことから、映像を見る際の弊害となるのである。

【0012】伝送されている画像の皮膚領域におけるひずみを低減させる、幾つかの技術では、場面に存する人の顔、手、その他の皮膚領域を含む映像の内容についての、質に関する情報を抽出することに注目している。これは、より少ないデータ圧縮成分を用いて、そのような認識された領域を符号化するためである。従って、これらの認識された領域は、より大きな値のビット速度（ビット/秒）を用いて、符号化され、伝送される。そこで、当該映像が復号された際、そのような認識された領域が、ひずみを受けた特徴を含む程度は少なくなるわけである。

【0013】ある技術では、映像イメージのシーケンスを、対称な形状を求めてサーチ（検索）する。ここで、対称な形状とは、対称軸について同一のもの半分ずつに分割可能な形状として定義される。対称軸とは、対象（物）を等しい部分へと分割する線分のことである。対称な形状の例としては、正方形、円、楕円などがある。映像における対象（物）が、対称な形状を求めてサーチ（検索）される場合、映像に示された顔や頭の一部については認識可能である。対称に描かれた顔や頭は、典型的な場合、楕円の形状に近いものであり、目の間、鼻の中心を通過して、口の中間を横切って縦方向に位置する、対称軸を有している。

【0014】半楕円のそれぞれは、対称である。というのは、それぞれ、一つの日、鼻の半分、口の半分を含んでいるからである。しかしながら、画像において対称的に描かれている顔及び頭のみが、認識可能であって、側面（側面の輪郭で）みた際には（左側を向いた場合あるいは右側を向いた場合）、顔や頭の認識はできない。なぜなら、側面（側面の輪郭で）みた顔や頭は、対称軸を含んでいないからである。場面に存する人の手やその他の皮膚領域についても、同様に対称なものではないし、また、対称をベースとした技術を用いては認識不可能である。

【0015】別の技術では、例えば楕円形、矩形、三角

形といった、特定の幾何学的形状を求めて映像をサーチ（検索）するものがある。特定の幾何学的形状につき映像をサーチ（検索）することで、しばしば頭及び顔を探し出すことが可能となるが、それでもなお、場面に存する人の手やその他の皮膚領域を認識することはできない。というのは、そのような領域は、典型的な場合、特定された幾何学的形状で表されることがないからである。さらに、特定された幾何学的形状に近似するものでもない、部分的に遮られたところのある顔及び頭も、同様に認識可能ではない。

【0016】さらに別の技術では、映像のシーケンスは、頭、顔、手を含む皮膚領域を認識するのに、色（色相）を用いてサーチ（検索）される。色（色相）をベースにした認識は、特定された皮膚のトーン（色調）の集合を用いて、適合する皮膚の色を有する対象（物）を求めて、映像シーケンスをサーチ（検索）することに依存している。色（色相）をベースにした認識は、ある場面について手、顔、その他の皮膚領域を認識するに有用である場合もある一方で、そのような領域の他の多くの場合、認識されることができない。

【0017】なぜなら、すべての人が、皮膚の同じトーンを有しているということはないからである。さらに、映像シーケンスの多くの皮膚領域における色の変化についてもまた、検出不可能であろう。その理由としては、適合する皮膚領域をサーチ（検索）するのに、特定された皮膚のトーンについての集合を用いるため、色をベースとした技術では、背景の照明あるいはシェーディングによる変化といった、対象（物）の色への予見不可能な変化を補償することができないことがある。

【0018】

【発明が解決しようとする課題】このようなことから、場面に存する人の手、顔、及びその他の皮膚領域を認識する皮膚認識技術が、求められ続けているのである。

【0019】

【課題を解決するための手段】本発明は、映像における皮膚領域を認識するための皮膚領域検出器に向けられたものであり、例示的な実施例では、映像符号化/復号化（Codec）装置の映像符号器（ビデオ・コーダ）と結びつけて用いられている。皮膚領域検出器は、映像シーケンスにおけるすべての対象（物）の形状を最初に解析し、皮膚領域を含んでいる可能性のある対象（物）の位置を割り出すことで、映像フレームにおける皮膚領域を認識する。皮膚領域を含んでいる可能性のある対象（物）はさらに解析され、そのような対象（物）の画素が、皮膚領域に特徴的な信号エネルギーを有しているかを判断する。

【0020】ここで用いられた信号エネルギーという語は、映像信号における、特定された画素グループについての輝度（明るさ）パラメータの二乗の合計を示している。信号エネルギーは、2つの成分を含んでいる。即

ち、直流（DC）信号エネルギーと交流（AC）信号エネルギーである。さらに、皮膚領域に特徴的な信号エネルギーを有する画素を伴った対象（物）についての色パラメータが、サンプリング（標本化）され、対象（物）に対する皮膚のトーンの値についての範囲を決定する。そこで、解析された対象（物）についての皮膚のトーンの値で、このようにサンプリングされたものの範囲は、映像に含まれたすべてのトーンと比較される。

【0021】これは、皮膚のトーンの値で同一の値を有する他の領域を、当該映像シーケンスにおいて認識するようにするためである。皮膚領域に特徴的な信号エネルギーの判断及び形状解析を基に、対象（物）における皮膚領域の可能性のある部分を認識するは有効であるといえる。というのは、皮膚のトーンの値についての範囲を判断するため、そのように認識された対象（物）の色サンプリングを繰り返して行うことで、対象（物）における色の変化を自動的に補償し、映像シーケンスの内容（コンテンツ）に関する皮膚検出は動的なものとなるからである。

【0022】ここでの例示的な具体例では、皮膚領域検出器は統合されているが、符号器、復号器、符号化コントローラといったものを含む映像符号化/復号化（Codec）装置の他の構成要素部分とは独立して機能している。一つの実施例では、皮膚領域検出器は、入力映像信号と符号化コントローラの間に挿入されており、映像の符号化に先だって、映像シーケンスにおける皮膚領域の位置に関連した入力を提供している。

【0023】本発明の一つの例においては、皮膚領域検出器は、形状ロケータ（形状位置決め器）とトーン検出器を含んでいる。形状ロケータ（形状位置決め器）は、入力映像シーケンスを解析し、映像フレームにおけるすべての対象（物）のエッジ（端部）を認識し、そのようなエッジが、皮膚領域を含んでいる可能性のある形状の輪郭に近似しているかを判断する。形状ロケータ（形状位置決め器）は、皮膚領域を含んでいる可能性のある、ある形状を認識するようにうまくプログラムされている。例えば、人の顔はおおよそ楕円の形状を有していることから、形状ロケータ（形状位置決め器）は、映像信号において楕円形状をした対象（物）をサーチ（検索）するようプログラムされている。

【0024】映像フレーム全体は、あまりに大きすぎるため、全体にわたって解析することはできないため、入力映像シーケンスの映像フレームは、最初に幾つかの画像領域に区切られている方が有利である。各画像領域についてはさらに、隣接する画素に対する、画素強度の大きさにおける変化を元に、対象（物）のエッジが決定される。各画像領域に及ぼす、隣接する画素に対する画素強度の大きさにおける変化が、特定された大きさよりも大きい場合には、そのような画像領域の位置は、対象（物）のエッジあるいはエッジの一部を含むものとして

認識されるのである。

【0025】その後、認識されたエッジあるいは認識されたエッジの一部は、さらに解析され、対象(物)の輪郭を表している。そのようなエッジが、皮膚領域を含んでいる可能性のある形状に近似しているかを判断する。皮膚領域は、大抵の場合、人間の形状のより滑らかな曲線(例えば、首筋やあご先の曲線)により画定されていることから、角張った境界は、典型的な場合、皮膚領域を示しているものではない。

【0026】従って、より滑らかな人間の形状に関連している輪郭は、大抵、皮膚領域を含んでいる可能性のあるものとして選択される。例えば、楕円形は、人の顔や頭の形状に近似していることから、楕円形に近似している対象(物)の輪郭を認識するのに、映像シーケンスの解析を行うことで、当該映像シーケンスにおいて皮膚領域を含んでいる可能性のある位置をうまく決定することとなる。また、映像による会議の場合、典型的には、少なくとも任意の人間がカメラに面していることから、誰かが部屋に在室していることすれば、楕円形状が認識される可能性がある。

【0027】皮膚領域を含んでいる可能性のある対象(物)について、一旦、形状ロケータ(形状位置決め器)により位置が突き止められると、トーン検出器は、位置が定められた対象(物)の画素を検査し、そのような画素が皮膚領域に特徴的な信号エネルギーを有しているかを判断する。さらに、そのような認識された対象(物)について、皮膚のトーンについての範囲をサンプリングし、サンプリングされた皮膚のトーンについての範囲をフレーム全体におけるトーンと比較し、適合する皮膚のトーンをすべて決定する。本実施例においては、輝度パラメータの信号エネルギー成分(DC及びACエネルギーなる構成要素)は、離散コサイン変換(DCT)技術を用いてうまく決定される。

【0028】本発明の技術では、皮膚領域を含んでいる可能性があるものとして認識された対象(物)における、特定された画素グループについての信号エネルギーの離散コサイン変換(DCT)が演算される。その後、各画素のACエネルギーによる成分は、当該離散コサイン変換(DCT)から、各画素のDCエネルギーによる成分を引く(減ずる)ことで決定される。

【0029】各画素についてのACエネルギーによる成分の値を元に、当該画素が、皮膚領域に特徴的なAC信号エネルギーを有しているかについての判断がなされる。検証された画素についてのACエネルギーが、特定された値よりも小さい場合には、典型的には、そのような画素は皮膚に関する画素として認識される。その後、トーン検出器は、そのような認識された画素の色パラメータをサンプリングして、対象(物)の領域内に含まれる、皮膚のトーンを示している色パラメータの範囲を決定する。

【0030】トーン検出器によりサンプリングされた色パラメータは、有利なことに、クロミナンスパラメータ C_r 、 C_b となっている。ここで用いられているクロミナンスパラメータという語は、映像信号についての色差の値を示しており、ここで、 C_r とは、赤色の成分と映像信号の輝度パラメータ(Y)との間の差として定義され、 C_b とは、青色の成分と映像信号の輝度パラメータ(Y)との間の差として定義される。続いて、トーン検出器は、サンプリングされた対象(物)から認識された皮膚のトーンの値についての範囲を映像フレームの残りに対しての色パラメータと比較して、他の皮膚領域を認識する。

【0031】その後、本発明の皮膚領域検出器は、映像シーケンスの次のフレームを解析して、皮膚のトーンの値についての範囲を決定し、次の映像フレームにおける皮膚領域を認識する。皮膚領域検出器は、選択的には、映像シーケンスの一つのフレームで認識された、皮膚のトーンの値についての範囲を用いて、映像シーケンスの次のフレームにおける皮膚領域を認識することも行う。

【0032】皮膚領域検出器は、選択的には、人の顔あるいは頭の形状に近似する対象(物)を解析するのに、目-鼻-口(eyes-nose-mouth, ENM)領域検出器を含んで、目-鼻-口(eyes-nose-mouth, ENM)領域の位置を決定する。一実施例では、ENM検出器は、形状ロケータ(形状位置決め器)とトーン検出器の間に挿入され、ENM領域の位置を認識し、そのような領域はトーン検出器による解析のベースとして用いる。目-鼻-口(eyes-nose-mouth, ENM)領域検出器は、対称性をベースとした方法を利用して、人の顔あるいは頭の形状に近似している対象(物)内に位置したENM領域を認識するものである。目-鼻-口(eyes-nose-mouth, ENM)領域が認識されるのは有効であるといえる。

【0033】というのは、顔のそのような領域は、例えば、目の色についてのパラメータ、眉毛の色のパラメータ、唇の色のパラメータ、及び髪の色のパラメータといった、皮膚のトーンについてのパラメータより他の色パラメータはもちろん、皮膚の色についてのパラメータも含んでいるからである。また、目-鼻-口(eyes-nose-mouth, ENM)領域の認識は、演算上の複雑さを低減させるものである。なぜなら、皮膚のトーンについてのパラメータは、認識された対象(物)の僅かな領域からサンプリングされるからである。

【0034】

【発明の実施の形態】図1は、本発明の例示的な応用例を示しており、ここで、皮膚領域検出器12は、例えば、映像コーデック(codec、符号器/復号器)10と、映像符号化/復号化システムと結びつけて用いられている。映像コーデック(codec、符号器/復号器)10のような映像符号化/復号化システムは、主として、画像圧縮技術をベースとして映像シーケンスの符

号化及び復号化を行う、電話会議に関連する産業において利用されている。

【0035】映像シーケンスの符号化及び復号化にとって有用な、画像圧縮技術の例としては、ITU-T勧告H.263(「狭帯域通信チャンネル用映像符号化」)において記述された、離散コサイン変換(Discrete Cosine Transform, DCT)による方法がある。もちろん、本発明は、映像符号器/復号器(codec、コーデック)より他の、例えば、映画編集装置といった映像システムについても有用であることは理解されるべきである。実際、本発明は、デジタルカラー映像信号が入力である任意の装置において利用に供されるべく応用可能である。

【0036】本発明の一実施例は、図1に例示されている。これは、映像コーデック(codec、符号器/復号器)10内に配置された、皮膚領域検出器12(点線で囲まれた部分)を示している。皮膚領域検出器12は、統合されているものの、映像コーデック(codec、符号器/復号器)10の他の構成要素部分とは独立して機能している。例えば、映像コーデック(codec、符号器/復号器)10は、映像符号器22、映像復号器24、符号化コントローラ16といった付加的な構成要素部分を含んでいる。そのような構成要素部分については、映像コーデック(codec、符号器/復号器)10の動作に関する以下の説明と結びつけて議論されることとなる。

【0037】図2のブロック図でより詳細に示された、皮膚領域検出器12は、形状ロケータ(形状位置決め器)50及びトーン検出器56を含んでいる。形状ロケータ(形状位置決め器)50及びトーン検出器56により提示された機能は、ソフトウェアを実行させることの可能なハードウェアを含めた、共有ないしは専用ハードウェアを利用することを通じて選択的に提供される。例えば、形状ロケータ(形状位置決め器)50及びトーン検出器56の機能は、単一の共有プロセッサか、あるいは複数の個々のプロセッサによって選択的に提供される。

【0038】また、形状ロケータ(形状位置決め器)50及びトーン検出器56を表している、個々の機能的なブロックを用いていることは、ソフトウェアを実行可能なハードウェアに専ら限って示しているものと解されるべきでない。上で記述された機能的なブロックについての付加的な例示の実施例としては、例えば、AT&T社のDSP16あるいはDSP32Cといった、デジタルシグナルプロセッサ(DSP)のハードウェア、以下で論じられる動作を実行するソフトウェアを保存するための読み出し専用メモリ(ROM)、デジタルシグナルプロセッサ(DSP)の結果を保存するランダムアクセスメモリ(RAM)を含んでいる。

【0039】また、選択的に、汎用デジタルシグナルプロセッサ(DSP)回路と組み合わせられたカスタムV

LSI回路同様に、超大規模集積(VLSI)ハードウェアによる実施例も考えられる。そういった実施例の任意のもの、あるいはすべてのものは、形状ロケータ(形状位置決め器)50及びトーン検出器56と表記される機能的なブロックの意味の中に入るものと考えられる。

【0040】本発明は、映像シーケンスにおける皮膚領域を認識するものである。形状ロケータ(形状位置決め器)50は、映像フレームにおけるすべての対象(物)のエッジの認識、及び、そういったエッジの任意のものが所定の形状の輪郭に近似しているかについての判断を基にして、映像フレームにおいて皮膚領域と目される部分を最初に探し出す。所定の形状に対する近似を基にエッジの解析を行うことは、重要なことといえる。

【0041】というのは、皮膚領域を含んでいる可能性のある対象(物)は、認識される高い確率を有しているからである。例えば、幾つかの例においては、人の顔あるいは頭は、楕円形の形状にほとんど近似していることになろう。従って、楕円形を認識するべく映像フレームの解析を行うことで、一部の皮膚領域の位置割り出しについて高い確率を提供するといえる。

【0042】皮膚領域の可能性のあるものとして認識された対象(物)は、その後、トーン検出器56により解析されて、そのような対象(物)の任意のものについての画素が、皮膚領域に特徴的な信号エネルギーを有しているかを判断する。このための開示で用いられた、信号エネルギーなる語とは、映像信号における特定の画素グループについての、輝度(明るさ)パラメータの二乗の合計を示しており、直流(DC)信号エネルギーと交流(AC)信号エネルギーという2つのエネルギー成分を含んでいる。そこで、皮膚領域に特徴的な信号エネルギーを有している画素を伴った対象(物)の色パラメータがサンプリングされて、当該対象(物)に対する皮膚のトーン(色)の値についての範囲を決定する。

【0043】当該対象(物)に対する皮膚のトーンの値についての範囲はさらに、映像に含まれたすべてのトーンと比較される。これは、当該映像シーケンスにおいて、皮膚のトーンについて同一の値を有する他の領域を認識するようにするためである。信号エネルギーについての解析、さらに、皮膚のトーンの値についてのサンプリングがなされることをベースとして皮膚領域が認識される際には、皮膚検出は、映像シーケンスの内容(コンテンツ)に関して動的になされているといえる。というのは、認識された対象(物)に対する皮膚のトーンのサンプリングを行うことで、背景の照明あるいはシェーディングによる変化といった、対象(物)のトーンに対する予見不可能な変化を自動的に補償するからである。

【0044】形状ロケータ(形状位置決め器)50及びトーン検出器56の両方の構成要素部分は、図2を参照して、皮膚領域検出器12の動作の説明の一部として

以下で記述されている。対象(物)の画像に対応するフレームシーケンスを時間の関数として表現している入力映像信号26は、従来の映像カメラ(ビデオカメラ)

(示されていない)から形状ロケータ(形状位置決め器)50へと供給されている。従来の映像カメラ(ビデオカメラ)としては、例えば、シャープ株式会社により製造されている、View Camがある。

【0045】形状ロケータ(形状位置決め器)50は、入力映像信号26のフレームの少なくとも一つを解析して、フレーム内のすべての対象(物)のエッジを認識し、エッジあるいはエッジの一部が、皮膚領域を含んでいる可能性のある形状に近似しているかを判断する。皮膚領域を含んでいる可能性のある形状の例としては、楕円形、弧形、及び曲線等がある。ここでの開示において用いられた曲線という語とは、直線ではないエッジの一部を少なくとも有している形状のことを示している。

【0046】形状ロケータ(形状位置決め器)50の構成要素部分は、図3で例示されており、粗スキャナ100、精細(細密)スキャナ102、形状フィッター104、はもちろん、形状位置アプロセッサ94を含んでいる。形状フィッター104は、形状ロケータ(形状位置決め器)信号106を生成し、これは、トゥーン検出器56へと供給される。

【0047】形状位置アプロセッサ94は、映像領域を解析し、映像フレームに含まれた対象(物)のエッジを認識するよう機能する。これらの形状位置アプロセッサ94は、ダウンサンプリング118、フィルタ120、デシメータ122、エッジ検出器124、しきい値回路126を含む。図4で例示されるような、アプロセッシング回路を組み込んでいる。

【0048】時間に関するダウンサンプリング118は、入力映像信号26において利用可能な全フレーム数から、解析目的のため、少数のフレームのみを選択することによって、形状認識のために利用可能な映像信号のフレーム数を制限するよう機能する。例示的な具体例としては、入力映像信号26のような映像信号についての典型的なフレームレートとしては、およそ毎秒30フレーム(30fps)で近似され、この場合、連続する各フレームは、前のフレームと本質的に同一の情報を含んでいる。

【0049】連続する各フレームが本質的に同一の情報を含んでいることから、形状解析のためには、映像信号から少数のフレームのみを選択することで、演算上の複雑さを低減させることが有効といえる。そこで、本例に関しては、演算上の複雑さを低減させるため、ダウンサンプリングは、形状解析のため入力映像信号について全フレームの4分の1のフレームのみを選択するものとする。その結果、ダウンサンプリングは、形状ロケータ(形状位置決め器)50への入力としてのフレームレートを、毎秒およそ30フレーム(30fps)というレートから

毎秒およそ7.5フレーム(7.5fps)というレートへと低減させる。

【0050】フィルタ120は、典型的には、ダウンサンプリングされた映像フレームの空間フィルタリングを実行するための分離形フィルタであり、 $3 \times 3 \times 2 \times 40$ の画素サイズを有し、 π/c という逆数周波数(カットオフ周波数)を有している。ここでcとは、有利なことに、以下で論じられるデシメーション(分割)ファクターに等しいものである。典型的な場合、フィルタ120のようなフィルタは、周波数の範囲を画定する。

【0051】ダウンサンプリングされた入力映像信号26のような信号がフィルタ120へ供給されると、映像信号に含まれた周波数で、当該フィルタについて画定された周波数の範囲内にあるもののみが出力となる。映像信号に含まれた周波数で、当該フィルタについて画定された周波数の範囲外にあるものについては、抑制されることとなる。フィルタ120の例としては、有限インパルス応答(finite impulse response, FIR)フィルタ及び無限インパルス応答(infinite impulse response, IIR)フィルタがある。

【0052】フィルタを通された映像信号は、デシメータ122への入力となる。ここで、エッジ解析のため、映像フレームの水平方向及び垂直方向の広がり、所定のより小さいサイズを持つ画像領域へと分割される。例示的な具体例として、デシメータ122のようなデシメータが、 $c=8$ というデシメーションファクター(デシメーション因子)を有しており、かつ、映像フレームが $3 \times 3 \times 2 \times 40$ の画素サイズを有しているとするれば、映像フレームは、 $4 \times 5 \times 30$ の画素サイズを有する画像領域へと分割される。

【0053】エッジ検出器124は、映像フレームの分割された画像領域のそれぞれについてエッジ検出を実行し、対象(物)のエッジを求めてサーチ(検索)を行う。任意の映像フレームにおける対象(物)のエッジは、典型的には、隣接する各画素についての画素強度の大きさにおける変化により特徴付けられる。例えば、 3×3 の画素サイズの画像領域が、対象(物)のエッジを含んでいないとすれば、そのような画像領域を表している。隣接する各画素についての画素強度の大きさは、以下の行列A(数1)に示されるようにほとんど等しいものとなる。

【数1】

$$A = \begin{bmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{bmatrix}$$

【0054】一方、 3×3 の画素サイズの同様の画像領域が、対象(物)のエッジを含んでいるとすれば、そのような画像領域を表している。隣接する各画素についての画素強度の大きさは、以下の行列B(数2)に示され

るように、鋭い遷移(変化)を含むこととなる。

【数2】

$$B = \begin{bmatrix} 10 & 50 & 90 \\ 50 & 50 & 90 \\ 90 & 90 & 90 \end{bmatrix}$$

【0055】エッジ検出器124といった、エッジ検出器は、ソーベル(Sobel)演算子を含む技術を利用して、分割された画像領域について、例えば行列AあるいはBで示された、隣接する画素についての画素強度の大きさと、 δ_x および δ_y といった二次元ソーベル(Sobel)演算子の畳み込みの二乗を合計することで、対象(物)のエッジを認識する。例示的な具体例としては、ソーベル(Sobel)演算子技術を用いて、水平方向には δ_x という

$$\delta_{xA} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{bmatrix}$$

結果としての畳み込みは、部分的には以下(数5)で示すような結果を生成する。

$$\delta_{xA} = (-1 \times 11) + (0 \times 10) + (1 \times 10) + (-2 \times 10) + (0 \times 10) + (2 \times 10) + (-1 \times 10) \\ + (0 \times 10) + (1 \times 11) = 0$$

$$\delta_{xA} = (1 \times 11) + (2 \times 10) + (1 \times 10) + (0 \times 10) + (0 \times 10) + (0 \times 10) + (-1 \times 10) \\ + (-2 \times 10) + (-1 \times 11) = 0$$

【0057】ここでの大きさは、二次元において0に近似しているものである。一方、例えば行列Bにおいて示された、隣接する画素についての画素強度のように、対象(物)のエッジを含む画像領域における隣接する画素についての画素強度の大きさをもって、ソーベル(Sobel)

$$\delta_{yB} = (-1 \times 10) + (0 \times 50) + (1 \times 90) + (-2 \times 50) + (0 \times 50) + (2 \times 90) + (-1 \times 90) \\ + (0 \times 90) + (1 \times 90) = 160$$

$$\delta_{yB} = (1 \times 10) + (2 \times 50) + (1 \times 90) + (0 \times 50) + (0 \times 50) + (0 \times 90) + (-1 \times 90) \\ + (-2 \times 90) + (-1 \times 90) = -160$$

ここでの大きさは0に近似するものではない。例えば、上で記述したソーベル(Sobel)演算子技術を利用して、エッジ検出技術は、映像フレームについての分割された45×30の画素領域のそれぞれにつき実行される。【0058】さらに、しきい値回路126は、分割された45×30の各画素領域における画素を認識する。そのような認識された画素に0でない数値を割り当てる際には、隣接する画素についての、畳み込まれ、二乗され、合計された画素強度の大きさは、特定された値よりも大きなものとなる。隣接する画素についての、畳み込まれ、二乗され、合計された画素強度の大きさが、しきい値回路126の特定された値よりも小さい場合には、画素には0という数値が割り当てられる。

演算子、垂直方向には δ_y という演算子により二次元形式で表現された、ソーベル(Sobel)演算子が以下の数3のように記述されているとすれば、

【数3】

$$\delta_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \delta_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

【0056】これらの演算子は、例えば行列Aについての隣接する画素に対する画素強度のような、対象(物)のエッジを含んでいない画像領域における、隣接する画素についての画素強度の大きさをもって畳み込まれる。これは以下(数4)のように示される。

【数4】

$$\delta_{xA} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \begin{bmatrix} 11 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 11 \end{bmatrix}$$

【数5】

1)演算子が畳み込まれる場合には、結果としての畳み込みは、部分的には以下(数6)で示すような結果を生成する。

【数6】

【0059】引き続き、0でない画素の値に対応するエッジデータ信号128が、しきい値回路126により生成される。しきい値回路126のような、しきい値回路を組み込むことで、輪郭を付けられた皮膚領域でエッジでないものが、エッジとして誤って認識されることをうまく防いでいる。というのは、隣接する画素についての画素強度の大きさにおける値かな変化によって、典型的な場合、畳み込まれ、二乗され、合計された(画素強度の)大きさが、しきい値回路126の特定された値よりも小さいものとなるからである。

【0060】再度、図3を参照すると、形状位置プリプロセッサ94により生成されたエッジデータ信号128は、形状ロケータ(形状位置決め器)50の粗スギ

ャナー100への入力となる。粗スキャナー100は、形状位置プリプロセッサ94により供給されたエッジデータ信号128を $B \times B$ の画素サイズのブロックへとセグメント化する(切り出す)。例えば、 5×5 の画素サイズのブロックへとセグメント化する。ブロックにおける画素の少なくとも一つが、上で論じたように、0でない値を有している場合には、各ブロックは、さらに粗スキャナー100によりマーキングされる。

【0061】そこで、セグメント化された $B \times B$ のブロックのエリアは、例えば、左から右へ、上部から下部へとといったようにしてスキャンされ、マーキングされたブロックの隣接するランを求めてサーチ(検索)がなされる。マーキングされたブロックについての、そのような各ランについて、精細(細密)スキャニングと形状フィッティングが実行される。形状ロケータ(形状位置決め器)50の構成要素部分として粗スキャナー100を含めることは、利用されるシステムについての演算上の複雑さによって、選択の余地がある。

【0062】精細(細密)スキャナー102は、セグメント化され、マーキングされた $B \times B$ のブロックについての隣接するランそれぞれにおける画素を、例えば、左から右へ、上部から下部へとといったようにしてスキャンし、画素の各列において、0でない値を有している最初の画素と0でない値を有している最後の画素を検出す。各列について検出された、0でない最初と最後の画素は、それぞれ、座標 (x_{start}, y) 及び (x_{end}, y) とラベル付けされる。

【0063】形状フィッター104は、画素の各列について (x_{start}, y) 及び (x_{end}, y) とラベル付けされた座標をスキャンする。形状フィッター104のメモリに保存された、様々なサイズ及びアスペクト比を持った幾何学的形状で、皮膚領域を含んでいる可能性のあるものは、そこで、ラベル付けされた座標領域と比較される。これは、おおよそその形状の適合を判断するためである。所定の形状についてよく適合する組み合わせから、例えば、楕円形といったような皮膚領域を含む可能性のある形状の輪郭を判断すると、形状ロケータ(形状位置決め器)50は、よく適合した形状についての座標を基に、形状位置信号106を生成し、そのような形状位置信号106をトーン検出器56へと供給する。

【0064】対象(物)が皮膚領域を含んでいる可能性を示している境界を伴った、当該対象(物)の位置について形状ロケータ(形状位置決め器)50が認識すると、トーン検出器56は、そのような対象(物)が、皮膚領域に特徴的な信号エネルギーを含んでいるかについて解析を行うよう機能する。当該対象(物)が、皮膚領域に特徴的な信号エネルギーを含んでいる場合には、皮膚のトーンの値についての範囲を認識するため、トーン検出器56は、当該対象(物)の色パラメータをサンプリングする。さらにトーン検出器56は、皮膚のトー

ンの値についての認識された範囲を映像フレームの残りに対しての色パラメータと比較して、皮膚のトーンの同じ値を含む他の領域を認識する。

【0065】カラーデジタル映像信号は、赤色(R)、緑色(G)、青色(B)という成分を含んでおり、典型的には、標準YUVカラー映像フォーマットで利用可能である。ここで、Yとは、輝度パラメータを表しており、U及びVはクロミナンスパラメータを表している。輝度(Y)パラメータは映像の明るさを特徴付けており、一方、クロミナンス(U、V)パラメータは、2つの色差の値、 C_r 、 C_b を定義している。輝度(Y)パラメータ、色差の値、 C_r 、 C_b と3つの色の成分R、G、Bとの間の関係は、典型的には次のように表現される。

$$Y = 0.299R + 0.587G + 0.114B$$

$$C_r = R - Y$$

$$C_b = B - Y$$

【0066】本発明の一実施例では、図5で示されたように、トーン検出器56は、皮膚領域検出器200、 C_r ヒストグラム生成器201、 C_b ヒストグラム生成器203、 C_r 範囲検出器205、 C_b 範囲検出器207、トーン比較器209を含んでいる。

【0067】皮膚領域検出器200は、入力映像信号26を形状位置信号106と相互に関連付け、そこで、映像フレームにおいて認識された対象(物)は、形状ロケータ(形状位置決め器)50により $D \times D$ の画素ブロックへとセグメント化される。 $D=2$ のときには、皮膚領域検出器200は、認識された形状を2×2の画素ブロックへとうまくセグメント化する。これは、2×2の画素ブロックすべてについて、一つの C_r の値と一つの C_b の値を得るためはもちろん、各画素について一つの輝度パラメータを得るためである。

【0068】例示的な具体例として、図6は、画素300についての 4×4 のブロックを示している。輝度パラメータ(Y)301は、各画素300について存在している。一方、画素300についての 2×2 の各ブロックは、一つの C_r の値302と一つの C_b の値303を有しており、水平及び垂直の両方向における中間点に存在している。このように、 2×2 の画素ブロックそれぞれは、一つの C_r の値302と一つの C_b の値303はもちろん、4つの輝度(Y)パラメータ301を含んでいる。上で論じたように、認識された対象(物)について皮膚のトーンのサンプリングが実行される際には、一つの C_r の値のみと一つの C_b の値のみを含めるべく、そのようなセグメント化を行うことは重要であるといえる。

【0069】 $D \times D$ の各画素ブロックが皮膚領域に特徴的な信号エネルギーを有しているかを判断することで、皮膚領域検出器200は、認識された対象(物)周辺にある $D \times D$ の画素ブロックいずれが皮膚領域を表しているのかを解析するよう機能している。カラー映像信号の

輝度 (Y) パラメータは、2つの信号エネルギー成分を有している。すなわち、交流 (AC) エネルギー成分と直流 (DC) エネルギー成分である。皮膚領域の画素は、典型的には、特定されたしきい値エネルギー T_{en} よりも小さい値を伴った交流 (AC) エネルギー成分を有している。

【0070】本発明の実施例では、カラー映像信号の輝度 (Y) パラメータについての交流 (AC) エネルギー成分に関する演算に基づいて、皮膚領域が検出される。ITU-T勧告H. 263 (「狭帯域通信チャンネル用映像符号化」) において記述された、離散コサイン変

$$F(u, v) = C(u)C(v) \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f(n, m) \cos \frac{(2l+1)u}{4} \pi \cos \frac{(2j+1)v}{4} \pi$$

ここで $F(u, v)$ とは、離散コサイン変換 (DCT) の関数を表しており、 $C(u)$ 及び $C(v)$ は、以下のよう

$$C(\omega) = 1/\sqrt{2} \quad 1/2 \quad 1/2$$

$$C(\omega) = 1$$

これらは、 $D \times D$ の画素ブロックについての各画素位置 $F(u, v)$ につき合計される。

【0072】さらに、交流 (AC) 信号エネルギー $E(m, 1)$ は、以下の数8の式に示されているように、離散コサイン変換の関数 $F(u, v)$ の二乗から直流 (DC) 信号エネルギー $F_{n,1}(0, 0)$ の二乗を引くことで決定される。

【数8】

$$E_{n,1} = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F_{n,1}(u, v)^2 - F_{n,1}(0, 0)^2$$

そこで、交流 (AC) 信号エネルギー $E(m, 1)$ は、しきい値エネルギー T_{en} と比較される。 $D \times D$ の各画素ブロックについて、交流 (AC) 信号エネルギー $E(m, 1)$ が、あらかじめ選択されたしきい値エネルギー T_{en} よりも小さい場合には、画素領域は、以下に示されるように、皮膚領域として認識される。

$E(m, 1) < T_{en}$ 皮膚領域

$E(m, 1) \geq T_{en}$ 非皮膚領域

【0073】典型的には、 $D \times D$ の画素ブロックが、120,000より小さい交流 (AC) 信号エネルギーの値を有している際には、そのような画素ブロックは皮膚領域として認識される。皮膚領域を判断するのに輝度パラメータの信号エネルギー成分を利用するのは有効であるといえる。というのは、非皮膚領域は、皮膚領域よりも、より高い信号エネルギー成分を有する傾向があるからである。そのような非皮膚領域を認識し、色サンプリングプロセスからそれらの領域を取り除くことで、サンプリングされた画素の色が、実際に皮膚領域の画素となる確率が増加し、従って、サンプリングされるトーンの範囲の精度が改善されることになる。

【0074】 $D \times D$ の画素ブロックが、皮膚領域検出器200により皮膚領域として一旦認識されると、当該D

換 (Discrete Cosine Transform, DCT) 技術を含む方法が、輝度 (Y) パラメータの信号エネルギーを演算するのにも有用である。

【0071】例示的な具体例として、 $D \times D$ の各画素ブロックについての輝度パラメータに関する交流 (AC) エネルギー成分と直流 (DC) エネルギー成分は、以下に示されているように、数7の式から、各画素について離散コサイン変換 (DCT) の関数 $F(u, v)$ を最初に演算することで決定される。

【数7】

に定義される。

$$(\omega = 0)$$

$$(\omega = 1, 2, 3, \dots)$$

$D \times D$ の画素ブロックについての C_x の値と C_y の値は、それぞれ、 C_x ヒストグラム生成器201及び C_y ヒストグラム生成器203によりサンプリングされる。前に論じたように、 $D \times D$ の画素ブロックが、 2×2 の画素ブロックであるとすれば有効である。というのは、そのようなブロックは、一つの C_x の値の一つの C_y の値のみを含んでいるからである。そこで、 C_x ヒストグラム生成器201と C_y ヒストグラム生成器203の両方は、サンプリングされた C_x と C_y の値それぞれについてのヒストグラムを生成する。

【0075】一旦、 C_x のヒストグラムと C_y のヒストグラムが生成されると、サンプリングされた対象 (物) についての皮膚のトーンを表している色パラメータの範囲は、統計解析技術を用いて、 C_x 範囲検出器204及び C_y 範囲検出器207を用いて決定される。例えば、サンプリングされた $D \times D$ の各画素ブロックに対する、 C_x と C_y の値の平均値とモード値が、各データセットについて決定される。 C_x と C_y の値の平均値とモード値が、それぞれについて特定された距離 D_p 内にある際には、そのような C_x と C_y の値の平均値とモード値は、単一のピークを表しているものとして認識される。

【0076】その後、 $D \times D$ の各画素ブロックについて、画素の色パラメータが、単一のピークを表す C_x と C_y の値の平均値とモード値についての所定の距離、例えば、ある標準偏差、の範囲内にある場合には、当該画素の色パラメータは、皮膚のトーンの値についての範囲に含まれる。平均値とモード値が、特定された距離 D_p よりも大きい際には、 C_x と C_y の値の平均値とモード値は2つの個別なピークを表しているものとして認識される。 $D \times D$ の画素ブロックについての画素の色パラメータが、2つの個別なピークを表している、 C_x と C_y の値の平均値とモード値を伴っている場合には、それらは皮膚のトーンの値についての範囲に含まれない。

【0077】 C_x 範囲検出器205及び C_y 範囲検出器207それぞれにおいて生成される、 C_x と C_y の値の範囲を基に、トーン比較器209は、入力映像信号26の全フレームを解析し、同一のクロミンスパラメータを含む、他のすべての領域の位置を探し出す。そのような他の領域の位置が明らかにされると、皮膚領域の位置を示している皮膚情報信号211がトーン比較器209により生成される。

【0078】皮膚領域検出器12は、映像シーケンスの各フレームについて上述の解析を実行するか、あるいは、選択的には単一のフレームを解析し、さらに、トーン比較器209は、皮膚のトーンの値についての範囲を利用して、次の、幾つかの特定された数のフレームにおいて皮膚領域を認識する。

【0079】本発明の実施例においては、形状ロケータ（形状位置決め器）50により認識された対象（物）の輪郭が、楕円形とうまく適合しており、そのような形状が皮膚領域を含んでいることが確かめられる以前に、形状ロケータ（形状位置決め器）50により生成された形状位置信号106が、選択的に、図7で示されるように、目—鼻—口（eyes-nose-mouth, ENM）領域検出器52へと供給される。

【0080】目—鼻—口（eyes-nose-mouth, ENM）領域検出器52は、形状ロケータ（形状位置決め器）50から、うまく適合した楕円形の輪郭の座標を受け取り、図8に示されているように、当該楕円領域を矩形ウィンドウ60及び楕円領域62（矩形ウィンドウ60内に位置していない、楕円形の残りの部分を含んでいる）へとセグメント化する。ENM領域検出器52は、楕円形のパラメータを受け取り、矩形ウィンドウ60が、目、鼻、口の領域に対応する楕円領域を把握すべく配置されるように、それらを処理する。

【0081】ENM領域検出器52は、サーチ（検索）領域識別子108を用いて矩形ウィンドウ60の位置決めを行うサーチ（検索）領域を決定する。ここで、図8で示されているような楕円形の輪郭についての中心点（ x_g, y_g ）の座標は、矩形ウィンドウ60の位置決めのためのおおよそを得るために用いられる。ENM領域の中心点の位置決めを行うサーチ（検索）領域は、例えば 12×15 の画素といたったように、 $S \times T$ の画素サイズの矩形であり、楕円形の輪郭の長軸及び短軸について決まったサイズを有するようにうまく選択される。

【0082】ここで開示で用いられた長軸という語は、図8に関連して定義されたものであり、点 x_1 及び y_1 の間で楕円形を二分する線分のことを示している。ここで開示で用いられた短軸という語はまた、図8に関連して定義されたものであり、点 x_1 及び x_2 の間で楕円形を二分する線分のことを示している。例示的な具体例として、楕円形が、長軸に沿って50画素の長さ、短軸に沿って30画素の長さを持つものとする。短

形ウィンドウ60のサイズは、 25×15 の画素サイズを有するようにうまく選択される。このサイズは、長軸と短軸の両方に沿って楕円形の半分の長さに近似的であるものであり、当該形状の目—鼻—口領域について、もっとも考えられる配置をとらえているものである。

【0083】矩形ウィンドウ60が楕円形内に一旦位置を定められると、サーチ（検索）領域スキャナ110は、矩形ウィンドウを解析し、当該楕円形の目—鼻—口領域に関しての対称軸についての候補となる各位置を決定する。例えば、サーチ（検索）領域スキャナ110は、左から右といった手法で、長軸と並行に置かれた線分64を用いて、矩形ウィンドウ60内の縦の画素行それぞれを選択する。これは、鼻の中心を通り、口の半分を通過し、両目の間に位置した、対称軸を求めてサーチ（検索）するためである。

【0084】鼻の軸に関する対称軸が決定された後には、ENM領域検出器52は、矩形ウィンドウ60についての結果として得られる目—鼻—口領域の座標に対応するENM領域信号54を生成する。ENM信号54は、対象（物）の目—鼻—口領域の位置に対する座標をトーン検出器56へ通知し、そこで、そのような領域に含まれていない画素は、以降の色パラメータ解析からは取り除かれることになる。

【0085】目—鼻—口領域が認識されるのは有効といえる。というのは、そのような顔の領域は、例えば、目の色のパラメータ、眉毛の色のパラメータ、唇の色のパラメータ、髪の色のパラメータを含む、皮膚のトーンのパラメータより他の色パラメータはもちろん、皮膚の色のパラメータを含んでいるからである。目—鼻—口領域で皮膚の色のパラメータを認識することは、サンプリングされる色パラメータの範囲についての精度を改善することになる。なぜなら、ENM領域の認識は、皮膚領域の存在を強く示すものであるからである。また、ENM領域は、それが導かれている、適切に適合した楕円形よりも小さいものであることから、演算上の複雑さも有効に低減される。

【0086】目—鼻—口領域の検出はまた、主体が直接カメラをみていないという場合に、影響が及ぼされる。これは、例えば、映像による電話会議の状況においてしばしば生じることである。ENM領域検出器52はまた、入力映像についての目—鼻—口領域の検出も包含しており、この場合、主体はカメラに直接面と向かっていわけではなく、顔に髪がもかかっていたり、眼鏡をかけていたりする。

【0087】ENM領域検出器52は、鼻を経て口を横切る縦軸につき、顔の特徴についての代表的な対称性を活用している。ここで、対称軸は、図8で示されているように、画像の縦軸に関して、角 θ_1 で傾斜される。そのような傾斜された楕円形については、目—鼻—口領域の検出におけるロバストネス（強健さ）を提供するた

め、矩形ウィンドウ60は、ウィンドウの中心について別個の値の角度だけ回転される。角 θ 1は、 -10° から 10° の範囲内の値を有していると都合がよい。

【0088】皮膚領域検出器12は、選択的には、映像コーデック(codec、符号器/復号器)10のような映像符号器/復号器(コーデック、codec)と結びつけて用いられる。以下の説明は、図1で示されたような映像コーデック(codec、符号器/復号器)10の他の構成要素に関する皮膚領域検出器12の動作を論じるものである。映像コーデック(codec、符号器/復号器)10は、映像符号器22及び映像復号器24を含んでおり、ここで、映像符号器22は、符号化コントローラ16によりコントロールされている。

【0089】符号化操作のため、映像コーデック(codec、符号器/復号器)10は入力映像信号26を受け取る。これは皮膚領域検出器12と映像符号器22へと前送りされる。皮膚領域検出器12は、上述のように入力映像信号を解析し、皮膚領域の位置に関係した情報を符号化コントローラ16へと供給する。映像符号器22は、符号化コントローラ16のコントロールの下で入力映像信号を符号化し、符号化ビットストリーム出力30を生成する。

【0090】ここで、上述の皮膚領域検出器を用いて認識された皮膚領域は、そのように認識されていない領域よりも高いビット数をもって符号化されている。例えば、符号化コントローラ16のような符号化コントローラは、典型的な場合、あるしきい値(量子化ファクター)以上の値を有する、離散コサイン変換のデータ成分のみを符号化し、伝送する。例示的な具体例として、 16×16 の画素領域が、1から16の範囲の値を有するデータ成分を有しており、しきい値が8に選択されたものとする。すると、符号化コントローラは、8というしきい値よりも大きい値を有する、DCTデータ成分のみを符号化することになる。

【0091】しかしながら、本発明の実施例においては、皮膚領域を含むものとして認識されている映像信号の部分について、しきい値以下の値を有するデータ成分は、ここで、しきい値以上の値を有するデータ成分と同じに符号化される。結果として、皮膚領域として認識されている映像領域は、そのように認識されない領域よりも高いビット数をもって符号化される。一実施例では、映像符号器22は、ソース符号器32、映像多重符号器34、伝送バッファ36、伝送符号器38を用いて入力映像信号26を符号化し、符号化ビットストリーム出力30を生成する。

【0092】復号化操作については、映像コーデック(codec、符号器/復号器)10は、符号化ビットストリーム入力40を受け取る。映像復号器24は、映像信号出力70を生成するため、受信復号器42、受信バッファ44、映像多重復号器46、及びソース復号器4

8を用いて、符号化ビットストリーム入力40を復号する。

【0093】

【発明の効果】本発明により、映像における人の手、顔、及びその他の皮膚領域の認識を行う皮膚認識技術において、従来の技術にみられた問題点を解決した皮膚領域検出方法及びシステムが実現された。とりわけ、本発明は、映像を用いた電話会議システムといった、マルチメディア環境での応用を目指した通信システムへ適用されることが期される。

【図面の簡単な説明】

【図1】図1は、本発明の原理についての例示的な応用例を具体化している、映像符号器/復号器(コーデック、Codec)のブロック線図である。

【図2】図2は、本発明の皮膚領域検出器のブロック線図である。

【図3】図3は、図2の形状ロケータ(形状位置決め器)のブロック線図を示している。

【図4】図4は、図3の形状ロケータ(形状位置決め器)についてのアプリロセッサー回路のブロック線図である。

【図5】図5は、図2のトーン検出器のブロック線図を示している。

【図6】図6は、 4×4 の画素ブロックを例示している。

【図7】図7は、目-鼻-口(eyes-nose-mouth, ENM)領域検出器を含む、皮膚領域検出器のブロック線図を示している。

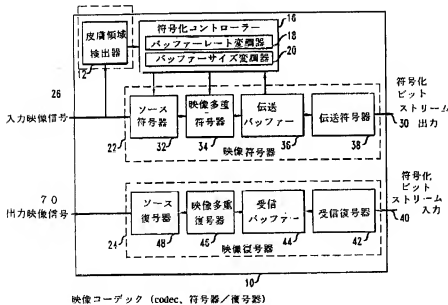
【図8】図8は、楕円形内に配置された矩形ウィンドウを例示している。

【符号の説明】

- 10 映像コーデック(codec、符号器/復号器)
- 12 皮膚領域検出器
- 16 符号化コントローラ
- 18 バッファレート変調器
- 20 バッファサイズ変調器
- 22 映像符号器
- 24 映像復号器
- 26 入力映像信号
- 30 符号化ビットストリーム出力
- 32 ソース符号器
- 34 映像多重符号器
- 36 伝送バッファ
- 38 伝送符号器
- 40 符号化ビットストリーム入力
- 42 受信復号器
- 44 受信バッファ
- 46 映像多重復号器
- 48 ソース復号器
- 50 形状ロケータ(形状位置決め器)

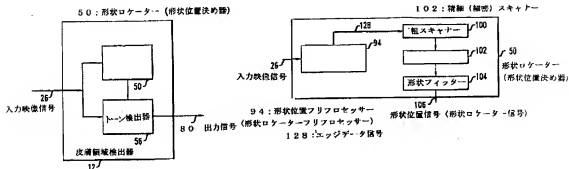
- | | | | |
|-----|------------------------------------|-----|-----------------|
| 52 | 目-鼻-口 (eyes-nose-mouth, ENM) 領域検出器 | 120 | フィルター |
| 54 | 目-鼻-口 (eyes-nose-mouth, ENM) 領域信号 | 122 | デシメータ |
| 56 | トーン検出器 | 124 | エッジ検出器 |
| 60 | 矩形ウィンドウ | 126 | しきい値回路 |
| 62 | 相補領域 | 128 | エッジデータ信号 |
| 70 | 出力映像信号 | 200 | 皮膚領域検出器 |
| 80 | 出力信号 | 201 | C_e ヒストグラム生成器 |
| 94 | 形状位置アプリアプロセッサ (形状ロケータアプリアプロセッサ) | 203 | C_b ヒストグラム生成器 |
| 100 | 粗スキャナ | 205 | C_r 範囲検出器 |
| 102 | 精細 (細密) スキャナ | 207 | C_b 範囲検出器 |
| 104 | 形状フィッター | 209 | トーン比較器 |
| 106 | 形状位置信号 (形状ロケータ信号) | 211 | 皮膚情報信号 (出力) |
| 108 | サーチ (検索) 領域識別子 | 300 | 4x4 の画素ブロック |
| 110 | サーチ (検索) 領域スキャナ | 301 | 輝度パラメータ (Y) |
| 118 | (時間) ダウンサンプラ | 302 | C_e の値 |
| | | 303 | C_b の値 |

【図1】

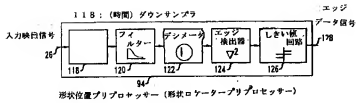


【図2】

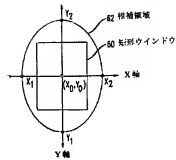
【図3】



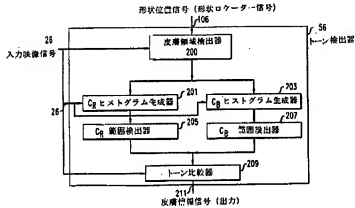
【図4】



【図8】



【図5】



【図6】

